PTO/SB/96 (6-98)
Approved for use through 09/30/2000. OMB 0651-0031
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(b)

Applicant/Patent Owner: KENDALL, Chad; CHOW, Joey; and NESBITT, Robert J.

Application No./Patent No.: Unassigned Filed/Issue Date: Herewith

Entitled: HIGH SPEED SEQUENCED MULTI-CHANNEL BUS

ALCATEL CANADA INC., a corporation,
(Name of Assignee) (Type of Assignee, e g., corporation, partnership, university, government agency, etc.)

states that it is:

1. [X] the assignee of the entire right, title, and interest; or

2. [ ] an assignee of an undivided part interest

in the patent application/patent identified above by virtue of either:

A. [X] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the Patent and Trademark Office at Reel ______, Frame ______, or for which a copy thereof is attached.

OR

B. [ ] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as shown below:

1. From:______________________________ To:______________________________
The document was recorded in the Patent and Trademark Office at
Reel ________, Frame ________, or for which a copy thereof is attached.

2. From:______________________________ To:______________________________
The document was recorded in the Patent and Trademark Office at
Reel ________, Frame ________, or for which a copy thereof is attached.

3. From:______________________________ To:______________________________
The document was recorded in the Patent and Trademark Office at
Reel ________, Frame ________, or for which a copy thereof is attached.

[ ] Additional documents in the chain of title are listed on a supplemental sheet.

[X] Copies of assignments or other documents in the chain of title are attached.
[**NOTE**: A separate copy (*i.e.,* the original assignment document or a true copy of the original document) must be submitted to Assignment Division in accordance with 37 CFR Part 3, if the assignment is to be recorded in the records of the PTO. See MPEP 302-302.8]

The undersigned (whose title is supplied below) is empowered to sign this statement on behalf of the assignee.

NOVEMBER 19, 2001
Date

Signature

HUBERT DE PESQUIDOUX
Typed or printed name

CEO, ALCATEL CANADA INC
Title

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

HIGH SPEED SEQUENCED MULTI-CHANNEL BUS

Technical field

[0001] This invention relates to telecommunication devices. In particular, the invention relates to methods and apparatus for transporting packets at high speed over a pin-limited interface while maintaining packet integrity and sequencing. The invention has application in devices such as switches and routers.

Background

[0002] A general problem in the telecommunications field is to convey data packets between different devices or parts of a device in an efficient manner. Data switches are an example of a type of device in which this problem can occur. Switches are used to selectively route data signals to their destinations.

[0003] A typical switch comprises a number of line cards which each provide an interface to one or more data lines. The data lines typically comprise optical fibers. When a packet is received at a line card, the packet is sent to a fabric interface card that interfaces to a switch fabric. The switch fabric determines an output data line on which the packet should be sent to reach its destination. This determination may be made, for example, on the basis of information in the packet's header. The switch fabric then routes the packet, by way of a fabric interface card, to the line card which is connected to the output data line. An example of a switch which has this general architecture is the ALCATEL ™ model 7670 switch.

[0004] Communication between the line cards and fabric interface cards is generally provided over a midplane. The midplane is typically limited in terms of the number of signal paths that it can maintain between different cards. Consequently, the midplane can present a bottleneck which interferes with reaching the goal of higher throughput.

There is continual pressure to increase the rate at which packets can be handled. Currently it is desired to support the OC-192 standard which operates at 10 Gbps data rates.

[0005] Various protocols exist which could be used to provide high speed data communications over a midplane. None of these currently available protocols are ideal. Available protocols include POS-PHY4 (Packet over SONET - Physical Layer 4) and FlexBus4™. A problem with POS-PHY4 is that it is difficult to implement in a cost-effective manner using ASICs (application specific integrated circuits) and FPGAs (field programmable gate arrays). FlexBus4 requires an interface which has an undesirably high pin-count. This increases the cost of providing switch hardware which uses the FlexBus4 protocol to carry data.

[0006] Lower-rate interfaces such as Utopia3 (universal test and operations physical interface for ATM) exist. Utopia3 provides a 32 bit bus operating at up to 100 MHz for data rates up to 3.2 Gbs. However, it has so far not been considered to be practical to provide increased bandwidth by using several such interfaces simultaneously because cells can get out of sequence.

[0007] There remains a need for cost-effective methods for transmitting data at high speed between devices over a pin-limited interface. There is currently a particular need for such methods capable of handling OC-192 data rates, which require an aggregate bandwidth of at least 12.8 Gps. In the future it will be desirable to accommodate higher data rates.

Summary of the Invention

[0008] This invention provides methods and apparatus for conveying data between points which are connected by an interface, particularly an interface which provides a limited number of signal conductors.

[0009] One aspect of the invention provides data transmission apparatus. The apparatus comprises a first transmit interface for transmitting a data stream comprising a sequence of fixed-size cells to a receiver. The first transmit interface comprises a first demultiplexer connected to receive the data stream and to split the data stream by delivering the cells in rotation into a plurality of N transmit channels so that each transmit channel carries every N^{th} cell; for each channel a data transmitting device connected to receive the cells of the transmit channel and to output the cells on one or more data connections to a receiver; and, a first transmit control circuit connected to the data transmitting devices, the transmit control circuit configured to cause the transmitting devices to output the cells in sequence with the commencement of transmission of cells on sequential transmit channels staggered in time relative to one another by a time difference ΔT. The transmitting devices may each comprise a serializer device and the data connections may comprise serial data connections.

[0010] Another aspect of the invention provides data transmission apparatus comprising: means for carrying a data stream comprising a sequence of cells in an order; demultiplexing means for assigning each of the cells of the data stream to one of a plurality of channels; transmitting means for transmitting the cells in each channel to a receiver; and, control means for commencing the transmission of individual cells to the receiver, in the order, at times staggered relative

to one another by a time difference ΔT. The transmitting means may comprise means for serially transmitting the cells in each channel to a receiver. The apparatus may comprise means for receiving a plurality of cells transmitted serially in a plurality of channels by another data transmitting device and means for determining an order of arrival of the plurality of cells.

[0011] Another aspect of the invention provides a telecommunications switch comprising a plurality of line cards, a switching fabric, a plurality of fabric interface cards connected to the switching fabric and a midplane providing a plurality of data lines connecting the line cards and the fabric interface cards. The switch comprises at least one bidirectional interface connecting one of the line cards and one of the fabric interface cards. The bidirectional interface carryies a first sequence of data cells in a data stream received at the line card in a first direction from the line card to the corresponding fabric interface card and a second sequence of data cells in a second direction from the fabric interface card to the line card. The bidirectional interface comprises: a first demultiplexer connected to receive the first data stream and to split the first data stream into a plurality of N first direction channels so that each first direction channel carries every N^{th} cell; for each first direction channel, a serializer device connected to receive the cells of the first direction channel and to output the cells as serial data on one or more serial data connections extending through the midplane to the fabric interface card; a first transmit control circuit connected to the serializer devices, the transmit control circuit configured to cause the serializer devices to output the cells in sequence order with the commencement of transmission of cells on different first direction channels staggered in time relative to one another by a time difference ΔT; a plurality of deserializer devices at the fabric interface card, the deserializer devices connected to receive and

deserialize the serial data on the serial data connections; a first direction receive control circuit connected to detect an order of arrival of cells on the serial data connections and to place the cells into a received data stream in the order of arrival; a second demultiplexer at the fabric interface card and connected to receive the second data stream and to split the second data stream into a plurality of N second direction channels so that each second direction channel carries every Nth cell; for each second direction channel a serializer device connected to receive the cells of the second direction channel and to output the cells as serial data on one or more serial data connections extending through the midplane to the line card; a second transmit control circuit connected to the serializer devices, the transmit control circuit configured to cause the serializer devices to output the cells in sequence order with the commencement of transmission of cells on different second direction channels staggered in time relative to one another by a time difference ΔT; a plurality of second deserializer devices at the line card, the deserializer devices connected to receive and deserialize the serial data on the serial data connections; and, a second direction receive control circuit connected to detect an order of arrival of cells on the serial data connections and to place the cells into a received data stream in the order of arrival.

[0012] The invention also provides a method for transmitting a data stream comprising a sequence of fixed-size cells to a receiver. The method comprises: assigning consecutive cells of the data stream into different ones of a plurality of channels; and, simultaneously transmitting to the receiver data on each of the channels while staggering transmission of consecutive ones of the cells in time relative to one another by a time difference ΔT. The method may comprise serializing the data of each channel before transmitting the data of the channel.

[0013] Still another aspect of the invention provides a method for transmitting a sequence of cells, in order, from a transmitting device to a receiving device. The method comprises: assigning each of the cells to one of a plurality of channels in rotation, each of the channels having a recurring cell transmit time, the cell transmit times for successive channels staggered relative to one another by amounts exceeding any inter-channel differences in skew and latency; in each channel, transmitting the cells in sequence to the receiving device over one or more serial data connections and commencing transmission of each cell only at the cell transmit time for that channel. The method may comprise receiving and deserializing the transmitted cells at a receiving device, and detecting an order of arrival of the cells at the receiving device.

[0014] Further features and advantages of the invention are described below.

Brief Description of the Drawings

[0015] In drawings which illustrate non-limiting embodiments of the invention:

Figure 1 is a block diagram of major components of a switch in which data is carried across a midplane according to the invention;

Figure 2 is a more detailed block diagram of a portion of the switch of Figure 1;

Figure 3 is a timing diagram showing a relationship between cell streams in multiple channels;

Figure 4 shows signals in a bidirectional interface according to one embodiment of the invention; and,

Figure 5 is a block diagram illustrating a bidirectional interface between a pair of cards according to the invention.

10171 - 04638960

<u>Description</u>

[0016] Throughout the following description, specific details are set forth in order to provide a more thorough understanding of the invention. However, the invention may be practiced without these particulars. In other instances, well known elements have not been shown or described in detail to avoid unnecessarily obscuring the invention. Accordingly, the specification and drawings are to be regarded in an illustrative, rather than a restrictive, sense.

[0017] Figure 1 shows a switch **10** which is used to illustrate this invention. The invention is not limited to use in switches but has application to other devices in which data packets must be carried between parts of a device, or between different devices. Switch **10** is connected to a number of optical fibers **12** each carrying an OC-192 data stream at 12.8 Gb/s. The invention is not limited to OC-192 data.

[0018] Switch **10** comprises a number of line cards **14** each associated with at least one optical fiber input **12**. Switch **10** also comprises a number of switch fabric interface cards (FICs) **16** and a switching fabric **18**. A midplane **20** permits data to be communicated back and forth between line cards **14** and FICs **16**. Each FIC provides an interface to switching fabric **18**.

[0019] As shown in Figure 2, data received from an incoming signal at optical fiber input **12** is received at an ingress chip **22** on line card **14**. The received data is placed on a wide bus **24** operating at a relatively slow speed. Bus **24** may comprise, for example, a 128-bit wide data portion operating at 100 MHz (for an aggregate throughput of 12.8 Gb/s). Bus **24** may comprise additional bits for control signals or additional header information. The use of a wide bus **24** permits bus **24**

to operate at a rate that is compatible with both ingress chip **22** and a FPGA **26**.

[0020] FPGA 26 may provide various functions including header translation. Because ingress chip 24 and FPGA 26 operate at relatively slow speeds they can be made with less expensive technology than would be required if bus 24 operated at a higher speed. This is particularly advantageous with respect to FPGA 26 since, while FPGAs capable of operating at clock frequencies significantly in excess of 100 MHz are becoming available, such FPGAs can be very expensive.

[0021] Midplane **20** typically is pin-limited (i.e. it has too few available signal conductors to simply extend bus **24** between a line card **14** and a FIC **16**). The term "midplane" is not limited to specific physical locations of signal conductors relative to any cards. The term midplane includes structures which are referred to as backplanes. Providing midplane **20** with enough signal conductors to permit signal paths in excess of 128 bits wide to be established between line cards **14** and FICs **16** is undesirably complicated and expensive.

[0022] The system of the invention breaks the data in bus **24** into a number of channels in each direction for transmission to and from FICs **16**. In the illustrated embodiment there are four channels labeled "A", "B", "C" and "D". For simplicity, Figure 2 only shows channels in the direction from line card **14** to FIC **16**. Figure 4 shows a possible configuration for one bidirectional channel across a midplane **20**. Data going in the direction from line card **14** to FIC **16** is designated "Rx" because it is received on the input **12** of a line card **14**. This direction may be called a "first direction". Data going in the direction from FIC **16** to line card **14** is designated "Tx" because it is destined to be

transmitted to some other device by line card **14**. This direction may be called a "second direction".

[0023] Data from bus **24** may be separated into channels by suitable logic on line card **14**. The logic for dividing data from bus **24** into channels may be equivalently implemented in fixed hardware or software (if sufficient speed can be achieved). The logic may be termed a demultiplexing means. The logic may be provided in FPGA **26**. Data in the channels is transmitted serially between line card **14** and FIC **16** and therefore requires a reduced number of data lines.

[0024] To accomplish this, line card **14** comprises a plurality of serializer devices **28** and FIC **16** comprises a corresponding plurality of deserializer devices **29**. In the illustrated embodiment, FPGA **26** comprises logic for dividing the data on bus **24** into four channels. FPGA **26** sends the data of each channel to a transmitting device for the channel. In the illustrated embodiment, each transmitting comprises a serializer device **28** on a bus **27**. The transmitting devices may be called "transmitting means". Each bus **27** may operate at the same frequency as bus **24**.

[0025] The data for each channel is routed from a serializer device **28** to a corresponding deserializer device **29** by way of a bus **30**. Bus **30** comprises a number of data lines which extend through midplane **20**. Serializer devices **28** could comprise, for example, model DS90CR483 serializer components available from National Semiconductor. Deserializer devices could comprise, for example, model DS90CR484 deserializer components available from National Semiconductor. A serializer device **28** and a deserializer device for data going in the opposite direction can conveniently be combined into a serializer/deserializer (SerDes) device.

[0026] Each channel carries a portion of the data from bus **24** together with various control signals. For example, there may be four channels. Each channel carries some of the data from bus **24**, a parity line, a start-of-cell signal, a control signal in the direction of data flow (CLAV - cell available), a control signal in the direction opposite to data flow (RXENB - receive enable or TXENB - transmit enable), and a clock signal.

[0027] A serializer device **28A** on line card **14** receives parallel data from a corresponding bus **27** and outputs that parallel data as one or more streams of serial data on bus **30** which connects to a corresponding FIC **16**. In a typical embodiment of the invention, each serializer device **28** receives a data stream 48 bits wide from bus **27** and outputs 8 streams of serial data. A clock signal at the frequency of bus **24** occupies a 9th data stream. Bus **30** operates at a higher rate than bus **27** so that throughput is maintained. In an example embodiment of the invention bus **30** operates at 700 MHz. Bus **30** connects to a corresponding deserializer device **29A** at its destination on FIC **16**.

[0028] Each data stream typically requires 2 data lines through midplane **20**. It can be seen that with this construction, all of the 12.8 Gb/s data from bus **24** can be transmitted across midplane **20** using only 36 pairs of data lines. This is a substantial reduction compared to the number of data lines in bus **24**. Data in the channels is recombined into a single data stream when it reaches its destination (in this case FIC **16**). At deserializer devices **29**, the data sent by serializer devices **28** is received and converted back into parallel data. This may be done by any suitable multiplexing means. The data is then placed onto a parallel data bus **32** which carries the data to a switch interface device **42**.

[0029] In the preferred embodiment, the data being transmitted comprises fixed-size data packets or “cells”. Each cell is delivered in one channel. The cells are assigned to channels in rotation. For example, where there are four channels A through D, the first cell is assigned to channel A, the second cell to channel B, the third cell to channel C, the fourth cell to channel D, the fifth cell to channel A, and so on. The cells can be assigned to channels in rotating order by a suitable demultiplexing means. Each of the N channels receives every N^{th} packet.

[0030] This arrangement can optionally be made to operate in dual modes. In a first mode, which is described above, a single data stream is separated into channels for transmission across midplane **20**. In another mode, ingress device **22** receives a number of separate lower-rate data streams and sends each separate stream across midplane **20** in a separate channel. For example, such a device could transmit a single stream of OC-192 data in its first mode and four streams of OC-48c data in its second mode. A dual mode device would have multiple inputs to ingress device **22** (or multiple ingress devices **22**).

[0031] As noted above, it is necessary to provide a mechanism which can cause the cells to be received in order at destination FIC **16**. The method of the invention maintains cell ordering by staggering the cell streams in each of the channels relative to one another. As shown in Figure 3, there is an interval ΔT between the time that transmission of a packet commences on one channel and the time that the transmission of the next packet commences on the next channel. Cells are only transmitted on ΔT boundaries. ΔT is chosen to be sufficiently large that packets are guaranteed to arrive at the destination deserializer devices **29** in order, and that this order can be correctly determined, despite any worst-case inter-channel variations in latency and skew. A control

circuit, which may be called a control means, controls each of the transmitting devices so that the transmission of each cell begins on a ΔT boundary. The control circuit may, for example, comprise a clock at intervals of $N\Delta T$ as shown in Figure 3.

[0032] Where the arrival of a cell is detected by sampling for a signal, such as a start-of-cell signal, at a specified clock rate, the start-of-cell signal will be detected within one cycle of the sampling clock. Where the sampling clock operates at 100MHz, the start-of-cell signal will be detected sometime within 10 ns after it is asserted. Where OC-192 data is being transmitted across a midplane in four channels and a 100 MHz clock is used to sample for the arrival of cells it is convenient to choose ΔT to be approximately 40 ns. This allows 10 ns for the detection of the start of a cell plus 30 ns to accommodate any inter-channel variations in latency and skew.

[0033] In normal operation, a cell is sent on one channel every 40ns. However, if there is no cell to send, or if one channel is not enabled, the transmitter should wait 40 ns before starting to send another cell. Cells should only be sent on ΔT boundaries. When a channel is not enabled, the transmitter should automatically try the channel again after a time ΔT has elapsed. In the alternative, the transmitter could proceed to the next channel after a time ΔT.

[0034] The reception of cells at FIC **16** is coordinated by a receiver control circuit **40**. As data is received by deserializer devices **29** it is placed in first in / first out (FIFO) buffers **34**. Receiver control circuit **40** places cells from buffers **34** onto bus **30** in the order that the cells are received. To do this, receiver control circuit **40** may determine the order in which start-of-cell signals, which indicate the arrival of a cell are received in the individual channels. In the illustrated

embodiment, receiver control circuit **40** comprises a small FIFO buffer **36**. When a deserializer device **29** receives a cell it sends a signal **37** which is stored in FIFO buffer **36**. The data in FIFO buffer **36** therefore indicates the order of arrival of cells at FIC **16**. Receiver control circuit **40** can sample FIFO **36** every ΔT (e.g. every 40 ns) to determine which of FIFOs **34** to service. Since cells may be dropped, receiver control circuit **40** only stores in FIFO **36** information regarding cells which have arrived and have been stored in a FIFO **34**. This may be implemented, for example, by permitting each channel to generate and send to FIFO **36** an end-of-cell signal **37** to receiver **40** which keeps track of the order of the cells and services FIFOs **34** in order.

[0035] In general, it is desirable to provide flow control signals for each channel. A receive-enable signal (RxEnb) is generated at receiver control circuit **40**. A serializer device **28** on line card **14** is inhibited from transmitting on a channel unless RxEnb is set for the channel. Since the RxEnb signals are traveling in a direction opposite to the flow of data in bus **30**, RxEnb signals are preferably multiplexed into data in channels going in the opposite direction (in this case, from FIC **16** to line card **14**). Figure 4 shows an example of signals in a bidirectional interface according to the invention. As shown in Figures 4 and 5, a TxEnb signal, which is multiplexed with the channel data, is used to control the transmission of data from FIC **16** to line card **14**.

[0036] Because of the latency associated with the transmission of data back and forth between line cards **14** and FICs **16**, a pipelining concept is used for the timing of the RxEnb signals controlling the delivery of cells in each channel. If there were no latency then, upon a FIFO **34** receiving the last cell it can hold, receiver **40** could deassert the RxEnb signal and thereby prevent additional cells from being transmitted. Serializer device **28** would react to the deassertion of the

RxEnb signal in time to defer sending a next cell on that channel. The latencies involved in the round-trip signal path between line cards **14** and FICs **16** make this impractical.

[0037] Accordingly, each FIFO **34** is made large enough to hold at least $\boldsymbol{P}$ cells (where $\boldsymbol{P}$ is an integer). When FIFO **34** is holding $\boldsymbol{P} - \boldsymbol{Q}$ cells (where $\boldsymbol{Q}$ is an integer in the range of $1 \le \boldsymbol{Q} \le (\boldsymbol{P}-1)$) then receiver **40** deasserts RxEnb for that channel. $\boldsymbol{Q}$ is selected so that even in the worst case latency through the system, the transmitter (e.g. serializer device **28**) will stop sending cells in time to prevent FIFO **34** from overflowing.

[0038] Figure 5 shows one channel of a bidirectional interface according to the invention. A first card (e.g. line card **14**) includes a first transmit interface and a first receive interface. The first transmit interface comprises a first demultiplexer, and a first data transmission device. In the illustrated embodiment the first data transmission device comprises a serializer device **28A**. Transmission of cells by serializer device **28A** is controlled by a first transmit control circuit.

[0039] The first receive interface comprises a deserializer device **29A'** a multiplexer, and a first receive control circuit. In the illustrated embodiment, the first transmit control circuit and the first receive control circuit are combined in a control circuit labeled "CONTROL".

[0040] In Figure 5, a second card (e.g. FIC **16**) has a second transmit interface and a second receive interface. The second receive interface may be connected to receive data from the first transmit interface. The second transmit interface may be connected to transmit data to the first receive interface.

09968646.112101

[0041] Where a component (e.g. an assembly, device, memory, etc.) is referred to above, unless otherwise indicated, reference to that component (including a reference to a "means") should be interpreted as a reference to any component which performs the function of the described component (i.e. is functionally equivalent to the described component), including components which are not structurally equivalent to the disclosed structure which performs the function in the illustrated exemplary embodiments of the invention. Where a step in a method is referred to above, unless otherwise indicated, reference to that step should be interpreted as a reference to any step which achieves the same result as the step (i.e. is functionally equivalent to the described step), including steps which achieve a stated result in different ways from those disclosed in the illustrated exemplary embodiments of the invention.

[0042] As will be apparent to those skilled in the art in the light of the foregoing disclosure, many alterations and modifications are possible in the practice of this invention without departing from the spirit or scope thereof. For example:

- The number of channels does not need to be four. More or fewer than four channels may be provided in each direction;
- The serializer devices and deserializer devices may be combined with other devices;
- The data in each channel does not need to be carried serially across midplane **20**. The data could be transmitted across midplane **20** as parallel data transmitted at a high enough rate that a reduced number of signal conductors are required in the midplane. The terms "transmitting devices" and "transmitting means" include serializer devices and devices for transmitting the data in each channel as parallel data;

10141245.046960

- The demultiplexing means may be incorporated in ingress device **22** or in an additional device.

Accordingly, the scope of the invention is to be construed in accordance with the substance defined by the following claims.